(a)

1) E C C C E E E E E E E E A A A D D D E E E E E E B B B C

2) D D D C D D D D E E E E E E E E A C C C A A A A D E E E

MRP $P_{i,j}$

	A	B	C	D	E
A	0	0.046	0.1668	0.4748	0.3124
B	0.2224	0	0.705	0.0606	0.0119
C	0.3365	0.1486	0	0.1375	0.3774
D	0.1664	0.1549	0.3599	0	0.3188
E	0.3104	0.1997	0.2449	0.245	0

FOMP $T_{i,j}$

	A	B	C	D	E
A	0.5169	0.0251	0.0636	0.2411	0.1533
B	0.0593	0.6674	0.2468	0.0244	0.0021
C	0.077	0.0404	0.7469	0.0374	0.0983
D	0.0349	0.0341	0.072	0.7946	0.0644
E	0.055	0.034	0.038	0.0389	0.8341

MRP λ

A	B	C	D	E
1	2	3	4	5

$\lambda=1$ $\lambda=2$ $\lambda=3$ $\lambda=4$ $\lambda=5$

(b)

5-g utilization (y-axis: 0.05 to 0.45)

Renewal process   1 order estimate

Figure 1.   (Colour online) Comparison of the MRP and the FOMP. (a) Examples of strings generated by the two processes. The first string (1) is generated by the MRP defined by transition table **P** and Poisson variances **λ** (with the corresponding distributions shown to the right). The FOMP transition table **T** is the maximum likelihood estimator of the transition table, generated from an extended version of string (1). Note that while the MRP **P** has zeros along the main diagonal (as repeated characters are defined by the Poisson vector **λ**), the FOMP **T** attempts to capture both the transitions between characters, and the character repeats, using a single matrix. The second string (2) is an example of a sequence of characters generated by **T**. (b) Statistical difference between sequences generated by the two processes. In a sequence of 10^4 characters, the MRP (left) generates approximately 10% of the $5^5 = 3125$ possible 5-gram combinations. In contrast, in sequences generated by the FOMP derived from the maximum likelihood estimator of the transition table, approximately 20% of the possible 5 gram appear.

ESM Figure 1. A graphical description of the differences between the RP and Markovian processes (reproduced from [8] Kershenbaum, A. 2013 Entropy rate as a measure of animal vocal complexity. Bioacoustics.)
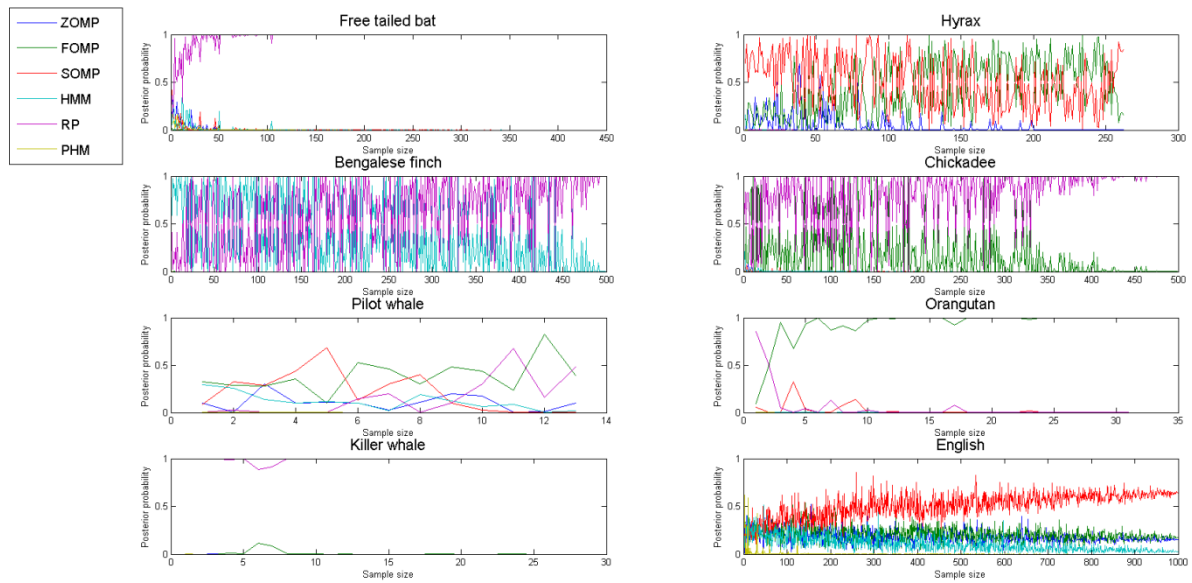
ESM Figure 2. Effect of sample size on optimal model classification. Each graph shows the results for a separate species data set. The y-axis shows the posterior probability of each of the tested models (legend on the left), for varying sample size. Sample sizes less than the full data set were generated by randomly selecting a sub-set of the vocal sequences. For each sample size, the average of 10 random repetitions were taken.

Note that for all the data sets except the pilot whale and the hyrax, a clear preferred model emerges by the full sample size.

ESM Table 1. Results of the Z-test for all models and all species. Numbers indicate the p-values when testing the null hypothesis that the sequences belong to the particular model. The model chosen by the Bayesian classifier as the best fit is shown in bold.

Species	Model type					
	ZOMP	FOMP	SOMP	HMM	RP	PHM
Free-tailed bat	0.900	0.012	0.929	0.227	**0.764**	0.820
Hyrax	0.000	**0.000**	0.070	0.000	0.050	0.000
Bengalese finch	0.000	0.000	0.000	0.108	**0.824**	0.000
Chickadee	0.000	0.632	0.065	0.401	**0.989**	0.000
Pilot whale	0.211	0.174	0.002	0.020	**0.586**	0.011
Orangutan	0.000	**0.026**	0.000	0.000	0.061	0.000
Killer whale	0.000	0.000	0.000	0.000	**0.646**	0.000
English	**0.971**	0.628	0.000	0.329	0.777	0.072